



# **There's an App for That**

## **Network Visualization: What and How**

**Stephanie Ritchie**

Lead Librarian, Customer Services Unit  
United States Department of Agriculture  
National Agricultural Library  
[stephanie.ritchie@usda.gov](mailto:stephanie.ritchie@usda.gov)

**Jessica Sigman**

Chemical Librarian  
ASRC Federal Professional Services  
[jsigman@asrcfederal.com](mailto:jsigman@asrcfederal.com)

The third in a series of columns around data visualizations by three librarians at the University of Maryland (and one collaborator from the U.S. Environmental Protection Agency).

### **Introduction**

Network visualizations provide “a useful way to visually represent the relationships in real or theoretical social, physical or biological systems.” ([Zoss 2018](#)) These relationships can be anything from social media connections, co-author partnerships, and Internet hyperlinks; to the electrical power grid or predator-prey relationships within food webs. Network visualizations reveal the connections (referred to as relationships, edges, links or lines) between things (referred to as entities, vertices, nodes or points). Each of the relationships or entities may have specific attributes that characterize the type of relationship or entity.

Network visualization can be used in any discipline but is most often employed in the social and biological sciences. Social scientists use it to explore relationships between people. Biological scientists use network visualization to explore relationships between organisms and within systems. In the library and information science field, network visualization has been used for bibliometric analysis of citation or co-authorship relationships, as well as in combination with keyword and text mining for topic analysis.

Visualization patterns reveal clusters in data, highlighting the most important differences or similarities. Visualization tools can display data to highlight comparative relationships between entities in ways that numerical tables do not. Additionally, network visualization allows for display of data beyond familiar bar and pie charts, to show not just singular data elements, but many data elements and the connections between them.

## Key Features of Network Visualization

On the surface, a network data visualization (sometimes referred to as a ‘node-link graph’) is made up of just a few basic elements: nodes (or vertices), and edges (or links). A node represents a single data point, while an edge represents the connection between two nodes. In network analysis, a user’s goal is to explore the relationship (edge) between data points (nodes), and the patterns of that relationship. Along with the nodes and edges themselves, those patterns are what emerge in a network data visualization.

Clusters are statistically and visually significant groupings of nodes in a spatial context. Paths are a series of edges connected through multiple nodes. Clusters emerge when there is an increasing amount of association between a set of nodes on a path, indicating a statistical relationship within a subset of the data. These features denote patterns within the dataset.

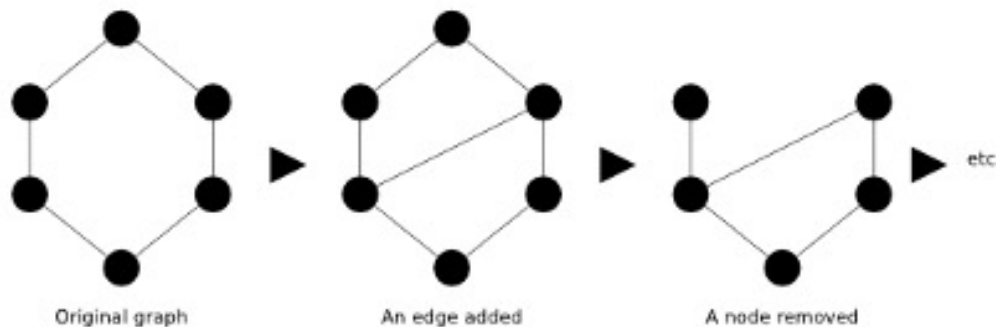


Figure 1: A basic network graph displaying nodes and edges. Image credit: [Wikimedia Commons](#)

The style of a network data visualization is critical in conveying the story the dataset tells. There are many style elements that can be employed in network visualization, but some of the most important include:

- Size: Node and edge size can denote the weight of the relationships and clusters found within the dataset
- Shape: Node shape can indicate significant subsets of data points or differentiate between data subsets that have a unique characteristic. Additionally, the use of shapes in edge representation can indicate spatial and temporal relationships between nodes/clusters.
- Color: Like size and shape, color plays an important role in differentiating between cluster groups, prominent pathways, and other significant patterns within the network. Ranking, a frequently used application of color, will apply a color range to a specific subset of the data that falls within a specific numerical range.
- Labeling: The addition of labels to a network visualization can play an important role in conveying specific additional details about the data story.

## Overview of Tools and Examples of Usage

The network visualization tools presented here are only a fraction of those available among a rapidly developing set of data visualization software available. All of the data visualization tools presented have a low barrier to use, i.e., do not require extensive coding knowledge or onerous training, although some do require installation of special programs. Most of these tools should be fairly straightforward for users familiar with structured data or simple web standards.

### *Cytoscape*

<http://cytoscape.org>

Cytoscape is a free open-source network visualization software originally designed for displaying molecular and biological interactions. In recent years, Cytoscape has been used in a broad range of disciplines as a tool for network analysis, and as such, has developed additional features for analysis and visualization.

Cytoscape offers a robust Desktop application as well as Cytoscape.js, an evolving JavaScript library which allows users to export visualizations to an interactive web format. Multiple suites of plugins related to network visualization are offered [Cytoscape App Store](#), including collections related to network generation, network analysis, cluster analysis, and network comparisons. Each analysis collection links to an associated Google Groups forum, where users can ask questions or search posts related to the tools in that collection.

Within the Cytoscape interface, users will find flexibility and interoperability with many current standards and workflows. Cytoscape offers easy backup and export options, through session files, and save formats. VizMapper, the built-in design tool, allows users to customize their network design and label style. The desktop application supports advanced filtering, browsing, and searching functionalities, as well as data in multiple languages. Current application plugins for Cytoscape support multiple data formats, fluid data integration, an expanded set of visualization packages and tools for analysis, and support for API development. Most importantly, Cytoscape continues to be free to use as an Open Access application.

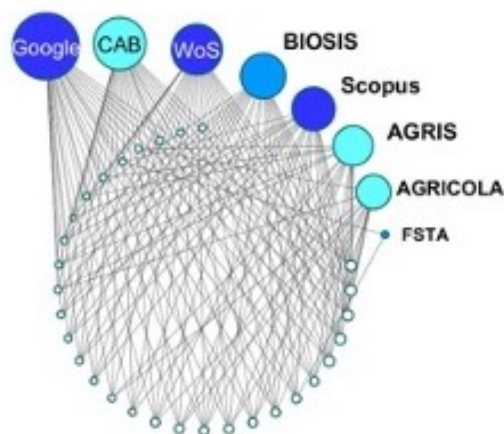


Figure 2: This circular network graph, created in Cytoscape, displays relationships between a set of citations and databases to illustrate the number of citations included in one database as compared to others. Bubbles increase in size as the number of connections increases. Databases with more connections are displayed in a gradient from left to right. Data available from [Ritchie et al., 2018](#).

## *Gephi*

<http://gephi.org>

Gephi is a free open source data exploration and visualization tool for large networks graphs which includes a 3D rendering functionality. Gephi offers several forms of network analysis including scientific and social, as well as association discovery, data carpentry, and poster creation. Gephi also supports API development and an online forum community for learning and troubleshooting. This has led to the development of many community generated plugins to meet the variable analysis and visualization needs of its community.

Gephi supports both small- and large-scale network visualization (e.g., Twitter mapping) and analysis, with the ability to manipulate datasets comprising up to 100,000 nodes and 1,000,000 edges. Their built-in data carpentry and analysis tool, Data Laboratory, aids in large scale data manipulation, analysis, and visualization. Additionally, Gephi includes unique forced layout algorithms, which allow the user to easily change the presentation of the network layout. Metric plugins and time series network analysis tools allow users to analyze dynamic evolution of a network over time.

The Gephi design interface includes many features to create visual presentations, including color and style changes, and ranking for optimal network appearance. In addition to supporting many file formats for import and export, Gephi provides extensive filtering, sorting, and search tools, as well as support in multiple languages.

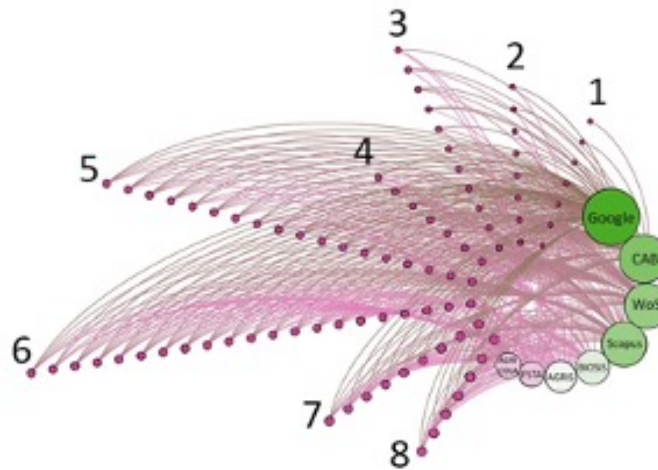


Figure 3: This radial axis network visualization, created in Gephi, shows relationships between a set of citations and databases to illustrate which databases contain the most content. Bubbles increase in size as the number of connections increases. Articles with the same number of connections are grouped in linear clusters around a central point. Data available from [Ritchie et al., 2018](#).

## *VOS Viewer*

<http://www.vosviewer>

VOSviewer is a free software tool for bibliometric data visualization including citation and co-authorship relationships as well as text mining of scientific literature. It can be used to create a “map” or visual display based on network, bibliographic or text data.

Sources for these data include Web of Science, Scopus, PubMed and Dimensions files, RIS and other citation management file types, and several direct APIs. Author and keyword relationships as well as occurrence may be visualized by selecting data elements and display preferences in terms of layout and scale.

Three types of map view offer flexibility in the level of detail projected in the visualization. The default display provides a bubble view using random colors for each cluster. Users have the option to overlay a gradient for a secondary data element (i.e., date published, number of citations, number of co-author connections, etc.). A density, or heat map, view highlights where a greater number of relationships exist within the data. A small selection of font and color options, as well as a dark background toggle, are available to customize the maps and improve accessibility.

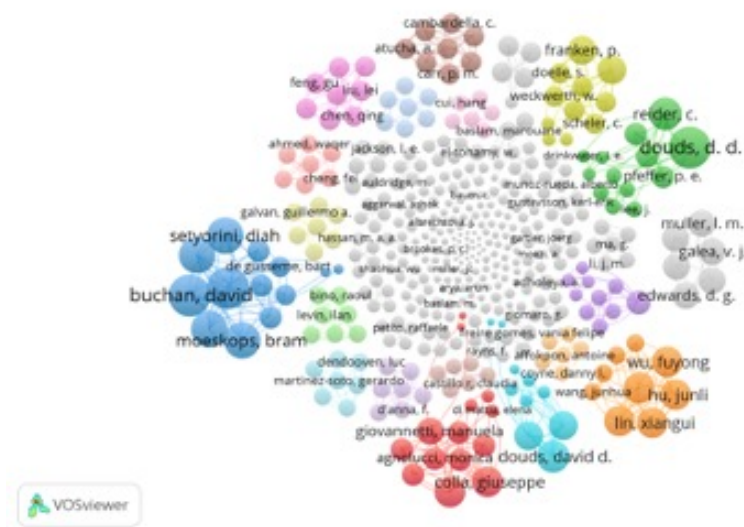


Figure 4: This clustered network visualization, created in VOSViewer, reveals co-author relationships for a topic search. Notice several distinct clusters have emerged and these clusters vary in size. Bubbles increase in size as the number of connections increase.

### *D3.js - Data-Driven Documents JavaScript library*

<https://d3js.org>

D3.js is a JavaScript library for visualizing data using code contributed by the D3.js community. Code for over 100 types of data visualizations ranging from standard bar and line graphs to maps and animations are available. This code is open and may be adapted to specific user and data needs. Common network visualization types available include force-directed graphs (clustered layout) and chord diagrams (circular layout).

Data may be transformed into interactive or static visual representations using a combination of scripts (JavaScript), web markup language (HTML), scalable graphics (SVG), and style sheets (CSS). Data formats compatible with D3 comprise comma separated values (csv), tab separated values (tsv), JavaScript object notation (json) and eXtensible markup language (xml) file formats. D3.js does not require specific software, rather a web server, browser, text editor, and access to the D3 library. Types of data visualizations and variants available to D3 users are not reliant upon a specific software capability or development, but on the ability of the user.

Novice JavaScript coders may build visualizations using existing code libraries while manipulating the layout, color, and other visual properties using website markup, style and graphic elements. Advanced JavaScript users may be compelled to share their edits, customizations, and improvements with the greater D3 community.

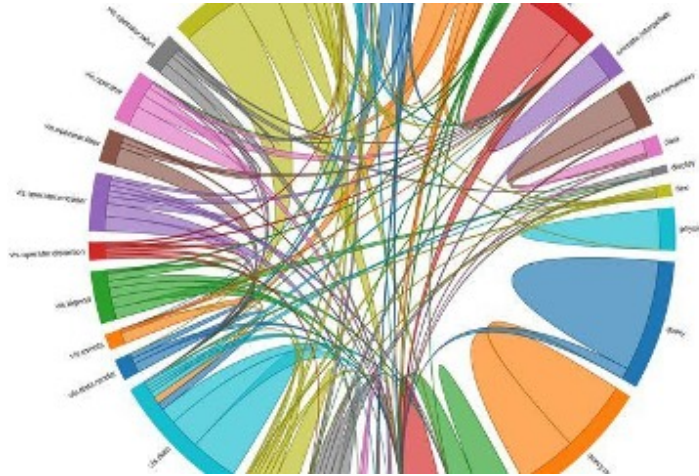


Figure 5: This chord diagram network visualization, created in d3js, displays relationships between and within software packages. Image credit: [d3js](#)

## Tips and Tricks

## Best practices

[Dunne and Shneiderman \(2009\)](#) have provided a set of guidelines for network visualizations:

- 1) Every node should be visible
- 2) Degree of node should be available
- 3) Links are available from source to destination
- 4) Clusters and outliers are identified

Network visualizations can easily become a random snarl of dots and lines, providing little useful information about the data. Dunne and Shneiderman's guidelines help ensure that network data can be clearly communicated. If the guidelines cannot be implemented, then the data should be revised to meet best practice standards, or a different data visualization method should be explored.

## Pitfalls

- 1) Garbage In, Garbage Out: It is important that you understand your data and the analysis you have performed on it. If there are errors in the data manipulation and analysis, your visualization will also contain errors.
- 2) With a few exceptions (Gephi), raw data often must be transformed into a format used by the network visualization tool. This may involve an additional step, and require tools such as Excel, R, and OpenRefine.

- 3) Your visualization is telling the story of your data. In order to do that, it must be clear, easily defined, legible, and intentional in its purpose. A confusing or messy visualization will hurt more than help.
- 4) Avoid bias in data representation: Bias can show up intentionally, and often, unintentionally. A good way to avoid bias is to have another researcher check your work. Additionally, thoroughly document all data and visualization manipulation you perform, and the reasoning behind it.

## Further Reading

**Dunne, C. & Shneiderman, B.** 2009. *Improving Graph Drawing Readability by Incorporating Readability Metrics: A Software Tool for Network Analysts*. [No. HCIL-2009-13]. p.9. Available from [https://www.cs.umd.edu/sites/default/files/scholarly\\_papers/CodyDunneUpdate\\_1.pdf](https://www.cs.umd.edu/sites/default/files/scholarly_papers/CodyDunneUpdate_1.pdf).

**Golbeck, J.** 2013. *Analyzing the Social Web*. Amsterdam (NL): Elsevier. Available from <https://www.elsevier.com/books/analyzing-the-social-web/golbeck/978-0-12-405531-5>.

**Mazza, R.** 2009. *Introduction to Information Visualization*. London (UK): Springer. DOI: [10.1007/978-1-84800-219-7](https://doi.org/10.1007/978-1-84800-219-7).

**Ritchie, S.M., Young, L.M. & Sigman, J.** 2018. A comparison of selected bibliographic database subject overlap for agricultural information. *Issues in Science and Technology Librarianship*. 89. DOI: [10.5062/F49Z9340](https://doi.org/10.5062/F49Z9340).

**Ward, M.O., Grinstein, G. & Keim, D.** 2010. Visualization techniques for trees, graphs, and networks. In: *Interactive Data Visualization: Foundations, Techniques, and Applications*. Boca Raton (FL): A K Peters/CRC Press. p. 271–290. Available from [http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/Chapter8\\_VisualizationTechniquesForTreesGraphsAndNetworks\\_271-290.pdf](http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/Chapter8_VisualizationTechniquesForTreesGraphsAndNetworks_271-290.pdf).

**Zoss, A., Maltese, A., Uzzo, S.M. & Börner, K.** 2018. Network visualization literacy: Novel approaches to measurement and instruction. In: Cramer, C. B., editor. *Network Science in Education: Transformational Approaches in Teaching and Learning*. Cham (CH): Springer. p. 169–187. DOI: [10.1007/978-3-319-77237-0\\_11](https://doi.org/10.1007/978-3-319-77237-0_11).



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).